International Journal of Coal Geology 146 (2015) 91–103

Contents lists available at ScienceDirect

International Journal of Coal Geology

journal homepage: www.elsevier.com/locate/ijcoalgeo

A metaproteomic approach for identifying proteins in anaerobic bioreactors converting coal to methane

Ji Zhang [a], Yanna Liang [a,*], Peter M. Yau [b], Rohit Pandey [c], Satya Harpalani [c]

[a] Department of Civil & Environmental Engineering, 1230 Lincoln Dr., Southern Illinois University Carbondale, Carbondale, IL 62901, USA
[b] Carver Biotechnology Center — Protein Sciences Facility, 600 S. Mathews Avenue MC-712, 315 Noyes Laboratory of Chemistry Box 1-1, University of Illinois, Urbana, IL 61801, USA
[c] Department of Mining and Mineral Resources Engineering, 1230 Lincoln Dr., Southern Illinois University Carbondale, Carbondale, IL 62901, USA

ARTICLE INFO

Article history:
Received 17 February 2015
Received in revised form 14 May 2015
Accepted 14 May 2015
Available online xxxx

Keywords:
Coal
Methane
Microbial consortium
Extracellular proteins
LC/MS

ABSTRACT

To understand the processes involved in bioconversion of coal to methane, a metaproteomic approach was taken to identify proteins in microcosms containing coal, standard medium and an adapted microbial community. Concentrated and dialyzed protein samples were subjected to further cleanup and trypsin digestion followed by mass spectrometric analysis. Searching the generated peaklists against domains of bacteria, archaea and fungi revealed 152 ± 1.4, 96.5 ± 2.1 and 38 ± 1.4 protein families, respectively. Proteins associated with bacteria were distributed among transporter and membrane proteins (33.1%), cellular metabolism (28.5%), substrate utilization/conversion (7.3%), oxidative stress (5.3%), cell movement (3.3%) and hypothetical proteins (22.5%). Among the total archaea proteins, 37.8% were for substrate utilization related to methane production, 27.6% were for cellular metabolism, 6.1% responded to stress, 5.1% were transporter and membrane proteins and 23.5% were those with unknown functions. Proteins produced by fungi fell in two groups: cell metabolisms (45.7%) and hypothetical proteins (54.3%). Based on key enzymes identified, a pathway for methanogenesis in the tested samples was proposed. This pathway illustrated methane production from four starting compounds, acetate, formate, methanol and CO_2. The proposed pathway will serve as a solid foundation for future effort aiming to increase methane yield from coal.

Published by Elsevier B.V.

1. Introduction

During recent years, considering the environmental drawbacks of generating electricity from coal combustion, converting coal to methane through biological processes has attracted significant attention (Fallgren et al., 2013; Wei et al., 2014). As a result, formation water samples collected from different coal seams have been evaluated in terms of the potential for producing methane. Several studies through constructing clone libraries or next generation pyrosequencing have been conducted for coals from the Powder River Basin (Ayers, 2002; Flores et al., 2008; Green et al., 2008; Ulrich and Bower, 2008), the San Juan Basin (Scott et al., 1994), the Illinois basin (Strąpoć et al., 2008), the Indio formation (Jones et al., 2010), the Alberta coalbeds in western Canada (Penner et al., 2010), the Jiuligang Formation in the Jingmen-Danyang basin in Hubei, China (Wei et al., 2014), the south Sydney Basin (Faiz and Hendry, 2006) and others listed in the review (Strąpoć et al., 2011). As a consequence, communities of fermentative and acetogenic bacteria and methane-releasing archaea have been identified in different subsurface environments. However, although the microbial distribution in a given place is known, the microbial functionality remains largely unclear.

Metabolic activities of a microbial community can be characterized by isotope analysis or through the analysis of methane production. From the perspective of molecular biology, cellular activities can also be revealed by the analysis of: 1) transcripts or metatranscriptome, the collective mRNA from all microorganisms in an ecosystem and 2) proteins or metaproteome, the collective proteins from all microbial species present in an ecosystem (Stokke et al., 2012). While the former provides insights into gene expression and activity, not all expressed genes will participate in certain pathways due to additional levels of cellular localization and regulation which occur at the protein level (Vanwonterghem et al., 2014). Thus, only results from the metaproteome study can give direct evidence of cellular metabolic activities at molecular levels.

Benefitted from the rapid development of mass spectrum instrumentation and bioinformatics software, metaproteomic analysis has been performed for various samples, such as: a mesophilic biogas-producing community fermenting straw and hay (Hanreich et al., 2013), a complex microbial community producing methane from agricultural waste and energy crops (Heyer et al., 2013), a microbiota in the phyllosphere and rhizosphere or rice (Knief et al., 2011), a microbial community from an anaerobic industrial-like wastewater treatment bioreactor (Abram et al., 2011), proteins present in the extracellular

* Corresponding author. Tel.: +1 618 453 7808; fax: +1 618 453 3044.
E-mail address: liangy@siu.edu (Y. Liang).

http://dx.doi.org/10.1016/j.coal.2015.05.006
0166-5162/Published by Elsevier B.V.

polymeric substances of active sludge flocs (Park et al., 2008), and an ANME (anaerobic methanotrophic archaea) community in marine cold seep sediments (Stokke et al., 2012). For communities degrading coal to methane, however, no such investigations have ever been conducted.

Recently, an original microbial community collected from a coalbed methane (CBM) well in the Illinois basin and an adapted consortium developed from it were studied through next-generation sequencing. Both the original and the adapted consortium contained bacterial and archaeal species and produced methane from coal in a laboratory setting (Zhang et al., 2015). To understand the functionality of the microbial community and the pathways leading to methane from coal, we aimed to identify proteins in anaerobic microcosms designed for bioconversion of coal to methane. Instead of using the traditional 2-dimension gel electrophoresis which is likely to result in biased results towards the most abundant proteins (Abram et al., 2011), we adopted the state-of-the-art proteomic approach to separate and identify target proteins. Based on the proteins detected, a pathway for methanogenesis was proposed here.

2. Materials and methods

2.1. Coal samples

Coal samples used in this study were the same as those investigated in another work (Zhang et al., 2015). Briefly, chunks of high volatile B bituminous coals were collected from Herrin Seam (No. 6) of the Illinois basin. The coals were ground and fragments that were retained between 40 and 100 mesh (0.15–0.425 mm) screen was stored in Ziploc bags and maintained in a humidity chamber to avoid water loss.

2.2. Methane production

A maintenance culture of the adapted consortium initially developed from the CBM community was established in our laboratory. To set up the duplicate microcosms, which were 100 mL serum bottles, the maintenance culture serving as the inoculum was added in a volume of 10% of the final total volume to 10 g fresh coal in 45 mL of a standard medium (Bonin and Boone, 2006). The two bottles were then closed with a butyl rubber stopper, sealed by an aluminum crimp and kept in dark at 28 °C.

To understand the bio-conversion process better, methane yields from different controls were also evaluated. These controls included: 1) both coal and the inoculum autoclaved. This was to test whether microbial contamination took place during microcosm cultivation; 2) coal with autoclaved inoculum. This was to evaluate whether microorganisms associated with coal can produce methane; and 3) the inoculum and medium only (without coal). This was to determine whether the consortium can generate methane from the supplemented medium. As stated above, for each condition, two replicates were established. All microcosms were maintained at 28 °C in the dark. At days 10, 20 and 30, samples from the headspace in each serum bottle were analyzed by gas chromatography (GC) as described previously (Zhang et al., 2015).

2.3. Protein identification

2.3.1. Sample preparation

Immediately after day 30, the microcosms were frozen at −20 °C. Upon use, the entire content in each microcosm was allowed to thaw first, followed by transferring to centrifugation tubes. The liquid portion after centrifuging the entire content at 4,000 g for 15 min was further

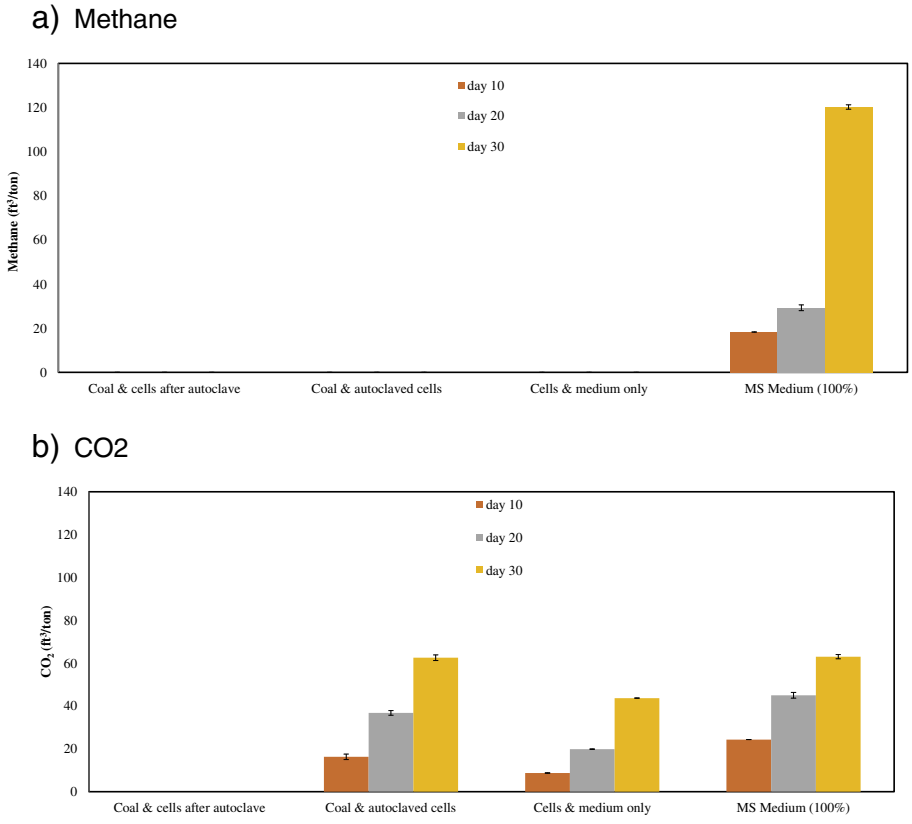

Fig. 1. Gas production from microcosms under different conditions. a: methane; b: CO_2.

Table 1
Summary of protein identification from different samples.

Samples	Bacteria			Archaea			Fungi		
	Total protein hits	Protein hits score ≥ 43	Average ± standard deviation	Total protein hits	Protein hits Score ≥ 43	Average ± standard deviation	Total protein hits	Protein hits score ≥ 43	Average ± standard deviation
#1	192	151	152 ± 1.4	164	98	96.5 ± 2.1	107	37	38 ± 1.4
#2	204	153		164	95		94	39	
#3	311	246	240.5 ± 7.8	197	85	64 ± 29.7	183	92	80 ± 17.0
#4	270	235		117	43		122	68	
#5	150	115		69	24		47	19	
In #1 + #2, not in #3, #4, and #5		39			37			10	

#1 and #2: Replicates of microcosms containing coal, the inoculum and the standard medium.
#3 and #4: Replicates of microcosms containing the inoculum and the standard medium, but without coal.
#5: One replicate containing coal and the standard medium, but without the inoculum. Another replicate was lost during sample preparation.

vacuum-filtered through 0.2 μm sterile filters. To the collected filtrate, a volume of 10 μl of Halt™ Protease Inhibitor Single-Use Cocktail EDTA-free (Pierce Biotechnology, Rockford, IL, USA) was added. The filtrate was then processed through Pierce Concentrators (9 K MWCO, 20 mL, Pierce) following the manufacturer recommended procedures. The concentrated protein samples were further dialyzed against distilled and deionized water three times. The dialyzed and concentrated samples were supplemented with 10 μl Protease Inhibitor to prevent protein degradation. Protein concentrations of these final samples were measured through using a BCA Protein Assay kit (Pierce) according to the manufacturer's protocol.

To prepare samples for protein identification, the samples were further cleaned by using Perfect Focus (G-Biosciences, St. Louis, MO, USA) according to the manufacturer's recommendation. Cleaned samples were digested with MSG-Trypsin (G-Biosciences) in 25 mM ammonium bicarbonate at a ratio of 1:10–1:50 (w/w) using a CEM Discover Microwave Digestor (Mathews, NC, USA) at 55 °C and maximum power of 60 W for 30 min. Digested peptides were lyophilized and resuspended in 5% acetonitrile plus 0.1% formic acid.

2.3.2. UPLC/MS

Ultra performance liquid chromatography (UPLC) was performed using a Thermo Dionex Ultimate RSLC3000 operating in nano mode at 300 nL/min with a gradient from water containing 0.1% formic acid to 100% acetonitrile + 0.1% formic acid in 200 min. The trap column used was a Thermo Acclaim PepMap 100 (100 μm × 2 cm) and the analytical column was a Thermo Acclaim PepMap RSLC (75 μm × 15 cm). The mass spectrometer used was the highly sensitive Thermo LTQ Velos Pro MS.

2.3.3. Data analysis

Xcalibur raw files were converted by Mascot Distiller into peaklists that were submitted to an in-house Mascot Server and searched against specific NCBI-NR protein databases for archaea, bacteria and fungi.

3. Results and discussions

3.1. Methane production

Headspace gas analysis of different microcosms revealed that (Fig. 1): 1) no CO_2 and CH_4 were observed in serum bottles with autoclaved coal and inoculum, which indicated that no microbial contamination took place during the 30-day cultivation period and autoclave adequately deactivated any microbial activities within coal and the inoculum, which was the acclimated microbial consortium; 2) increased CO_2, but no CH_4 was released with time from microcosms containing coal and the autoclaved inoculum, which demonstrated that the microbial strains associated with coal samples could degrade coal to CO_2. But these cells were not able to produce methane even though archaea strains similar to *Methanobrevibacter* sp. were present (Zhang et al., 2015); 3) increased CO_2, but no CH_4 was detected in serum bottles with the inoculum and medium only. For these setups, no coal was provided. Thus, the released CO_2 was from organic carbon in the medium provided. Grown on this nutrient solution, however, the adapted microbial consortium did not produce any CH_4 though species close to *Methanobacterium bryantii* and *Methanobrevibacter arboriphilus* were identified (Zhang et al., 2015). Therefore, all methane observed in our experiments was from coal itself but not from any nutrients provided. Nutrient, such as yeast extract was speculated to be a source for methane production during the first 72 h of cultivation (Green et al., 2008). But in this study, we proved that yeast extract and peptone did not lead to methane release; and 5) with the presence of coal and the inoculum, the standard medium gave a methane yield of 120.0 ft^3/ton in 30 days, which was similar to 111 ft^3/ton detected previously in 20 days (Zhang et al., 2015). These calculations assumed that the powdered coal samples we used were uniform in composition. Although the overall yield was close, the methane production rate of 4.0 ft^3/ton/day was lower than 5.6 ft^3/ton/day we observed before. The reason was that, for this experiment, we used coals that had particle sizes between 40 and 100 mesh, coarser than those used in previous experiments, which were <40 mesh. This is in agreement with previous report that finer coals lead to higher methane production rate even for lignite (Harding et al., 1993).

3.2. Protein identification

The conventional procedure for identifying proteins in metaproteomes generally has two steps: protein separation by 2-D polyacrylamide gel electrophoresis (PAGE) and protein identification by LC/MS/MS. Although this approach has its own advantages, in particular, related to allowing visual comparison of protein up- or down regulation, it suffers from the fact that only abundant proteins with high spot density on the gel can be accurately excised and identified. In addition to this drawback, mass spectrometer with low sensitivity hinders proper protein detection. As a result, some successful metaproteomic studies have reported identification of very limited number of proteins, such as, 17 (Hanreich et al., 2012; Heyer et al., 2013), 18 (Abram et al., 2011) and 36 (Hanreich et al., 2013). In the current study, however, the improved ion optics of the Thermo LTQ Velos Pro MS together with the nano-UPLC separation enabled us to achieve superior sensitivity and resolution for protein identification without going through gel separation. The great capability of this MS has been demonstrated by identifying 158 proteins in chicken egg white proteome (Mann and Mann, 2011) and other numerous studies.

For this investigation, processed protein samples from three sets of microcosms were analyzed by UPLC/MS. These three sets were: 1) two replicates which contained coal, the standard medium and the inoculum; 2) two microcosms which comprised the standard medium

Table 2
List of statistically valid proteins identical to those in bacteria. Highlighted are those that are present in samples #1 and #2, but not in samples #3, #4 and #5.

Number	Family	Accession	Score	Mass (kDa)	emPAI	Protein function	Related species
Transporter and membrane proteins (50)							
1	5	gi|504408682	239	24301	2.37	Outer membrane protein W	*Pseudomonas stutzeri*
2	6	gi|506253858	227	64900	0.58	ABC transporter substrate–binding protein	*Desulfomicrobium baculatum*
3	15	gi|503373989	130	38735	0.55	Membrane protein	*Sphaerochaeta globosa*
4	16	gi|503373018	120	40391	0.52	ABC transporter substrate–binding protein	*Sphaerochaeta globosa*
5	32	gi|503505202	91	35448	0.43	ABC transporter substrate–binding protein	*Sphaerochaeta coccoides*
6	14	gi|504409833	148	46123	0.32	Porin	*Pseudomonas stutzeri*
7	79	gi|584456509	64	18454	0.26	Putative membrane protein	*Clostridium sp.* M2/40
8	49	gi|503371944	78	39402	0.24	Membrane protein	*Sphaerochaeta globosa*
9	9	gi|652927745	56	46966	0.2	Porin	*Desulfovibrio alaskensis*
10	47	gi|653125520	80	23862	0.19	Membrane protein	*Chryseobacterium sp.* UNC8MFCol
11	145	gi|499687451	45	26953	0.17	Amino acid ABC transporter substrate–binding protein	*Desulfovibrio alaskensis*
12	48	gi|518368793	79	58230	0.16	Molecular chaperone GroEL	*Proteiniphilum acetatigenes*
13	20	gi|499687845	114	30583	0.15	Amino acid ABC transporter substrate–binding protein	*Desulfovibrio alaskensis*
14	27	gi|495986360	95	33313	0.14	C4–dicarboxylate ABC transporter substrate–binding protein	*Synergistes sp.* 3_1_syn1
15	142	gi|654714408	48	38483	0.12	Spermidine/putrescine ABC transporter ATP–binding protein	*Bradyrhizobium japonicum*
16	97	gi|488493334	61	37723	0.12	Phosphonate ABC transporter phosphate–binding periplasmic component	*Grimontia sp.* AK16
17	121	gi|506253299	54	36994	0.12	Phosphonate ABC transporter substrate–binding protein	*Desulfomicrobium baculatum*
18	129	gi|503372034	52	36984	0.12	Sugar ABC transporter substrate–binding protein	*Sphaerochaeta globosa*
19	55	gi|504229310	76	36613	0.12	Amino acid ABC transporter substrate–binding protein	*Rhodospirillum photometricum*
20	41	gi|499686540	84	37217	0.12	C4–dicarboxylate ABC transporter	*Desulfovibrio alaskensis*
21	24	gi|503371580	104	36683	0.12	Membrane protein	*Sphaerochaeta globosa*
22	37	gi|488746454	87	38420	0.12	Membrane protein	*Treponema denticola*
23	134	gi|500472186	50	40832	0.11	ABC transporter substrate–binding protein	*Geobacter uraniireducens*
24	89	gi|491901690	62	40070	0.11	ABC transporter substrate–binding protein	*Dethiosulfovibrio peptidovorans*
25	42	gi|502493564	84	41708	0.11	Branched–chain amino acid ABC transporter substrate–binding protein	*Desulfomicrobium baculatum*
26	56	gi|494357177	76	39192	0.11	C4–dicarboxylate ABC transporter	*Thiocapsa marina*
27	144	gi|488793309	47	41038	0.11	MFS transporter	*Treponema saccharophilum*
28	22	gi|610423070	108	124276	0.11	Membrane protein	*Draconibacterium orientale*
29	72	gi|502620476	66	41113	0.11	Membrane protein	*Sulfurospirillum deleyianum*
30	52	gi|495986920	78	45699	0.1	Amino acid ABC transporter substrate–binding protein	*Synergistes sp.* 3_1_syn1
31	92	gi|517601834	62	43869	0.1	Major facilitator transporter	*Arthrobacter sp.* 162MFSha1.1
32	78	gi|490645924	64	44864	0.1	Membrane protein	*Arcobacter butzleri*
33	103	gi|489379122	59	47964	0.09	Porin	*Pseudomonas stutzeri*
34	85	gi|663179244	63	58623	0.08	ABC transporter ATP–binding protein	*Streptomyces griseoluteus*
35	96	gi|567409584	61	58327	0.08	Molecular chaperone GroEL	*Tannerella sp.* oral taxon BU063 isolate Cell 2
36	70	gi|503373212	66	61987	0.07	ABC transporter substrate–binding protein	*Sphaerochaeta globosa*
37	106	gi|499686471	58	60943	0.07	ABC transporter periplasmic protein	*Desulfovibrio alaskensis*
38	125	gi|446938309	53	67348	0.07	Multidrug ABC transporter	*Bacillus cereus*
39	51	gi|492478790	78	64216	0.07	Starch–binding protein	*Parabacteroides distasonis*
40	128	gi|545061074	53	61344	0.07	Bacterial extracellular solute–binding s, 5 Middle family protein	*Clostridium bifermentans*
41	13	gi|492469922	78	122385	0.07	TonB–linked outer membrane protein	*Parabacteroides distasonis*
42	29	gi|492741854	95	120966	0.07	SusC/RagA family TonB–linked outer membrane protein	*Bacteroides massiliensis*
43	62	gi|499124738	72	67183	0.07	Membrane protein, putative	*Cyclobacteriaceae bacterium* AK24
44	117	gi|610423071	55	61144	0.07	Membrane protein	*Draconibacterium orientale*
45	140	gi|518368794	49	120326	0.07	TonB–dependent receptor	*Proteiniphilum acetatigenes*
46	87	gi|491597882	63	63367	0.07	Von Willebrand factor A	*Saccharomonospora cyanea*
47	44	gi|495426344	82	111998	0.04	Collagen–binding protein	*Parabacteroides johnsonii*
48	114	gi|546549875	56	107494	0.04	Putative Penicillin–binding protein	*Clostridium chauvoei*
49	90	gi|494400363	62	115346	0.04	TonB–dependent receptor	*Bacteroides cellulosilyticus*
50	65	gi|654954551	70	218980	0.02	Metallophosphoesterase	*Bacillus sp.* J13

Table 2 (*continued*)

Cellular metabolism (43)							
51	2	gi|652797147	368	49208	1.38	Glutamate dehydrogenase	*Clostridium viride*
52	63	gi|545045990	71	6259	0.91	HxlR-like helix-turn-helix family protein	*Peptoclostridium difficile*
53	46	gi|653244188	80	9128	0.57	50S ribosomal protein L27	*Prevotella brevis*
54	122	gi|157829716	54	9308	0.55	Chain A, Apo-biotin carboxyl carrier protein	*Escherichia coli* BL21
55	7	gi|503373500	212	38364	0.39	LacI family transcriptional regulator	*Sphaerochaeta globosa*
56	34	gi|40965242	87	14622	0.33	Elongation factor Tu	*Acidithiobacillus ferrooxidans*
57	112	gi|654360090	56	15398	0.31	Plasmid stability protein	*Rhizobium leguminosarum bv. phaseoli* CCGM1
58	23	gi|151302245	106	68259	0.28	AprA, partial	*Desulfobotulus sapovorans*
59	137	gi|671532764	49	19391	0.24	MarR family transcriptional regulator	*Streptomyces sp.* NRRL F-5123
60	94	gi|495986366	61	22402	0.21	TetR family transcriptional regulator	*Rhodococcus sp.* AW25M09
61	77	gi|490522537	64	22060	0.21	Heat shock protein GrpE	*Cronobacter sakazakii*
62	93	gi|550971460	61	22565	0.21	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	*Rhodopseudomonas sp.* B29
63	135	gi|499648068	50	26308	0.17	3-alpha-hydroxysteroid dehydrogenase	*Pseudoalteromonas haloplanktis*
64	39	gi|506254199	85	57739	0.16	Cytochrome C	*Desulfomicrobium baculatum*
65	26	gi|504035097	96	31437	0.14	LacI family transcriptional regulator	*Sphaerochaeta pleomorpha*
66	95	gi|497240667	61	31500	0.14	ATPase, histidine kinase/DNA gyrase B/HSP90-like protein	*Oscillatoriales cyanobacterium* JSC-12
67	120	gi|493631975	55	31685	0.14	Pyridoxal biosynthesis protein	*Thermanaerovibrio velox*
68	45	gi|547881037	82	32386	0.14	D-3-phosphoglycerate dehydrogenase	*Bacteroides sp.* CAG:770
69	61	gi|515084228	73	35847	0.13	Multispecies: transketolase	*Pseudomonas*
70	138	gi|558634743	49	35349	0.13	Biotin attachment protein	*Sporolactobacillus laevolacticus*
71	84	gi|18034225	63	35044	0.13	Adenosine-5'-phosphosulfate reductase alpha subunit, partial	*Desulfobulbus elongatus*
72	53	gi|490630658	76	41872	0.11	ISXO2-like transposase domain protein	*Leptospira weilii*
73	76	gi|573582856	64	41707	0.11	30S ribosomal protein S1	*Viridibacillus arenosi* FSL R5-213
74	132	gi|547475109	51	39312	0.11	Heat-inducible transcription repressor hrcA	*Firmicutes bacterium* CAG:536
75	28	gi|501110705	95	44965	0.1	Glutamate dehydrogenase	*Alkaliphilus oremlandii*
76	148	gi|635640571	44	45882	0.1	Arginine deiminase	*Cellulomonas sp.* KRMCY2
77	143	gi|505326012	48	44541	0.1	Probable carbamoyltransferase YgeW	*Coprococcus catus*
78	43	gi|506255585	84	46665	0.1	Sulfate adenylyltransferase	*Desulfomicrobium baculatum*
79	25	gi|518369253	104	48846	0.09	Glutamate dehydrogenase	*Proteiniphilum acetatigenes*
80	57	gi|666990850	75	49034	0.09	Glutamate dehydrogenase	*Clostridium sulfidigenes*
81	10	gi|492756187	180	49224	0.09	Multispecies: glutamate dehydrogenase	*Blautia*
82	147	gi|500468731	45	48001	0.09	N-acetylglucosamine-1-phosphate uridyltransferase	*Synechococcus sp.* RCC307
83	82	gi|646359166	63	46987	0.09	Folylpolyglutamate synthase	*Sphingomonas sp.* JGI 0001002-A17
84	17	gi|648635267	120	105809	0.08	Peptidase M16	*Proteiniphilum acetatigenes*
85	149	gi|495696384	44	55071	0.08	Coproporphyrinogen dehydrogenase HemZ	*Clostridium sp. Maddingley* MBC34-26
86	111	gi|493409168	56	61294	0.07	Fis family transcriptional regulator	*Chlorobium ferrooxidans*
87	127	gi|499685943	53	59964	0.07	Cytochrome C	*Desulfovibrio alaskensis*
88	18	gi|494833642	116	65701	0.07	ATP synthase subunit A	*Bacteroides plebeius*
89	58	gi|71042030	73	61784	0.07	Chain A, Phosphoenolpyruvate carboxykinase	*Actinobacillus succinogenes*
90	64	gi|489072089	71	81582	0.05	Patatin	*Chryseobacterium gleum*
91	124	gi|497936449	54	158340	0.03	DNA-directed RNA polymerase subunit beta'	*Myroides injenensis*
92	131	gi|494739126	51	162698	0.03	DNA polymerase III PolC	*Listeria fleischmannii*
93	141	gi|565876778	48	255532	0.02	Malonyl CoA-acyl carrier protein transacylase	*Paenibacillus sp.* JCM 10914
Substrate Utilization/Conversion (11)							
94	12	gi|499687298	165	53956	0.48	Iron hydrogenase	*Desulfovibrio alaskensis*
95	75	gi|167541422	64	22625	0.45	Methyl co-enzyme A reductase	*Uncultured bacterium*
96	108	gi|530330964	57	13397	0.36	Dissimilatory sulfite reductase beta subunit, partial	*Uncultured sulfate-reducing bacterium*
97	38	gi|309320752	85	33993	0.28	Chain A, putative lactate dehydrogenase	*Francisella Tularensis*

(*continued on next page*)

Table 2 (*continued*)

98	73	gi|518367747	65	35751	0.27	Malate dehydrogenase	*Proteiniphilum acetatigenes*
99	30	gi|518369542	95	34248	0.13	Glucokinase	*Proteiniphilum acetatigenes*
100	115	gi|490973191	56	34998	0.13	Phosphotransacetylase	*Anaerococcus prevotii*
101	104	gi|68500061	58	42010	0.11	Dissimilatory (bi)sulfite reductase alpha	*Olavius ilvae Delta 1 endosymbiont*
102	54	gi|344189466	76	43250	0.1	Chain B, Desulforubidin	*Desulfomicrobium Norvegicum*
103	74	gi|494941697	65	53838	0.08	Rhamnulokinase	*Cronobacter condimenti*
104	107	gi|5542158	57	55070	0.08	Chain L, Ni-fe-se Hydrogenase	*Desulfomicrobium Baculatum*
105	67	gi|503505024	70	96860	0.05	Glycoside hydrolase	*Sphaerochaeta coccoides*
Oxidative stress (8)							
106	19	gi|654478265	115	15109	1.29	Rubrerythrin	*Haliea salexigens*
107	4	gi|518370610	251	54189	0.6	Oxidoreductase	*Proteiniphilum acetatigenes*
108	31	gi|490645984	91	11515	0.43	Thioredoxin	*Arcobacter butzleri*
109	35	gi|648635887	87	21476	0.22	Superoxide dismutase	*Proteiniphilum acetatigenes*
110	50	gi|496439571	78	21303	0.22	Superoxide dismutase	*Thiorhodovibrio sp.* 970
111	105	gi|657673854	58	21770	0.21	Superoxide dismutase	*Dehalococcoidia bacterium* SCGC AB-539-J10
112	100	gi|501520656	60	22167	0.21	Peroxidase	*Geobacter bemidjiensis*
113	133	gi|494880819	51	42294	0.11	FAD dependent oxidoreductase	*Rhizobium sp.* PDO1-076
Cell movement (5)							
114	1	gi|510831444	396	51398	0.39	Flagellin protein	*Bacillus nealsonii*
115	11	gi|544697115	179	28703	0.16	Flagellin	*Clostridium sordellii*
116	91	gi|495882958	62	34421	0.13	Chemotaxis protein CheW	*Alishewanella*
117	80	gi|499955912	64	48425	0.09	Flagellin	*Shewanella frigidimarina*
118	98	gi|3098305	61	48627	0.09	Flagellin, partial	*Pseudomonas stutzeri*
Functions unknown (34)							
119	3	gi|545055966	299	30398	0.74	Hypothetical protein	*Clostridium bifermentans*
120	119	gi|545590999	55	7324	0.74	Uncharacterized protein	*Phascolarctobacterium succinatutens*
121	21	gi|652797730	109	42194	0.66	Hypothetical protein	*Clostridium viride*
122	136	gi|518072057	50	10746	0.47	Hypothetical protein	*Bacillus massilioanorexius*
123	110	gi|517838952	56	11009	0.46	Hypothetical protein	*Deinococcus aquatilis*
124	69	gi|152064343	68	12625	0.39	Protein containing DUF820	*Beggiatoa sp.* PS
125	113	gi|495412546	56	15168	0.32	Hypothetical protein	*Pseudoalteromonas sp.* BSi20495
126	151	gi|511083040	44	17544	0.27	Hypothetical protein	*Oscillibacter sp.* 1-3
127	71	gi|406888153	63	36860	0.26	Hypothetical protein ACD_75C02217G0002	*Uncultured bacterium*
129	33	gi|547265169	89	20537	0.23	Putative uncharacterized protein	*Acetobacter sp.* CAG:977
130	8	gi|518368298	183	42906	0.22	Hypothetical protein	*Proteiniphilum acetatigenes*
132	118	gi|545321708	55	23234	0.2	Hypothetical protein	*Actinomadura madurae*
133	83	gi|406883987	63	24734	0.19	Hypothetical protein ACD_77C00322G0006	*uncultured bacterium*
134	116	gi|655514895	56	26660	0.17	Hypothetical protein	*Prevotella sp.* HUN102
135	126	gi|506250092	53	26211	0.17	Hypothetical protein	*Leptotrichia buccalis*
136	60	gi|491906203	73	28081	0.16	Hypothetical protein	*Massilia timonae*
137	40	gi|488641239	84	32692	0.14	MULTISPECIES: hypothetical protein	*Clostridiales*
138	101	gi|547735447	59	36918	0.12	Putative uncharacterized protein	*Prevotella sp.* CAG:1320
139	123	gi|515892800	54	40618	0.11	Hypothetical protein	*Cyanobacterium PCC 7702*
140	66	gi|557400908	70	44753	0.1	Hypothetical protein, partial	*Uncultured Thiohalocapsa sp.* PB-PSB1
141	81	gi|630770890	63	44309	0.1	Hypothetical protein HY2_08090	*Hyphomonas sp.* T16B2
142	99	gi|504715444	60	46137	0.1	Hypothetical protein	*Desulfosporosinus meridiei*
143	130	gi|503373802	52	43787	0.1	Hypothetical protein	*Sphaerochaeta globosa*
144	139	gi|516287447	49	46639	0.1	Hypothetical protein	*Paenibacillus sp.* PAMC 26794
145	109	gi|518370777	57	50113	0.09	Hypothetical protein	*Proteiniphilum acetatigenes*
146	146	gi|518367724	45	49186	0.09	Hypothetical protein	*Proteiniphilum acetatigenes*
147	86	gi|521258007	63	53788	0.08	Hypothetical protein	*Psychrobacter sp.* G
148	88	gi|489087334	62	55919	0.08	Hypothetical protein	*Sphingobacterium spiritivorum*
149	102	gi|655124331	59	58387	0.08	Hypothetical protein	*Desulfonatronum lacustre*
150	150	gi|497065916	44	55553	0.08	Multispecies: hypothetical protein	*Fischerella*
151	68	gi|494075306	70	64355	0.07	Hypothetical protein	*Bermanella marisrubri*

Table 3
List of statistically valid proteins identical to those in archaea. Highlighted are those that are present in samples #1 and #2, but not in samples #3, #4 and #5.

Number	Family	Accession	Score	Mass (kDa)	emPAI	Protein function	Related species
Substrate utilizatioin related to methane production (37)							
1	1	gi|499343651	761	28090	5.09	Methyl–coenzyme M reductase	*Methanosarcina mazei*
2	2	gi|499343654	617	45430	1.8	Methyl–coenzyme M reductase	*Methanosarcina mazei*
3	8	gi|499330238	232	24419	1.37	Sulfite reductase	*Methanosarcina acetivorans*
4	3	gi|499342910	547	35104	1.33	Phosphotransacetylase	*Methanosarcina mazei*
5	20	gi|501688740	125	30332	0.75	Methylene tetrahydromethanopterin dehydrogenase	*Methanosphaerula palustris*
6	5	gi|499333927	299	62035	0.73	Methyl–coenzyme M reductase	*Methanosarcina acetivorans*
7	11	gi|499329835	175	34090	0.64	Methyltransferase Methanol–5–hydroxybenzimidazolylcobamide	*Methanosarcina acetivorans*
8	7	gi|499331135	246	28590	0.56	methyltransferase Methanol:5–hydroxybenzimidazolylcobamide	*Methanosarcina acetivorans*
9	10	gi|12751300	191	49899	0.41	methyltransferase MtaG	*Methanosarcina acetivorans*
10	17	gi|503093719	137	50583	0.4	Coenzyme F420 hydrogenase subunit alpha	*Methanoplanus petrolearius*
11	40	gi|499331310	68	16352	0.29	Flavodoxin	*Methanosarcina acetivorans*
12	24	gi|499342590	117	36447	0.26	Methylcobamide:CoM methyltransferase	*Methanosarcina mazei*
13	37	gi|500168793	69	46004	0.2	Methyl–coenzyme M reductase	*Methanoculleus marisnigri*
14	28	gi|2494436	104	51003	0.18	Acetyl–CoA decarbonylase/synthase complex subunit gamma	*Methanosarcina thermophila*
15	12	gi|182637450	165	50335	0.18	Monomethylamine methyltransferase MtmB	*Methanosarcina mazei Go1*
16	41	gi|499625345	67	27958	0.16	Methanol–5–hydroxybenzimidazolylcobamide	*Methanosarcina barkeri*
17	89	gi|499627806	45	27910	0.16	methyltransferase Methanol–5–hydroxybenzimidazolylcobamide methyltransferase	*Methanosarcina barkeri*
18	63	gi|500015040	50	28797	0.16	Methyl–coenzyme M reductase	*Methanosaeta thermophila*
19	16	gi|399513604	144	28419	0.16	Methyl coenzyme M reductase subunit A, partial	*uncultured archaeon*
20	33	gi|399513638	89	28350	0.16	Methyl coenzyme M reductase subunit A, partial	*uncultured archaeon*
21	23	gi|499768486	118	34862	0.13	Methylenetetrahydromethanopterin reductase	*Methanospirillum hungatei*
22	91	gi|503095289	45	35271	0.13	Methylenetetrahydromethanopterin reductase	*Methanoplanus petrolearius*
23	43	gi|490139462	64	76391	0.12	Formate dehydrogenase subunit alpha	*Methanofollis liminatans*
24	69	gi|496360291	48	42124	0.11	Mandelate racemase	*Metallosphaera yellowstonensis*
25	25	gi|490181023	112	45780	0.1	Methyl–coenzyme M reductase	*Methanoplanus limicola*
26	42	gi|584720	65	44309	0.1	Acetate kinase Methanol–5–hydroxybenzimidazolylcobamide	*Methanosarcina thermophila*
27	55	gi|499333779	54	50182	0.09	methyltransferase	*Methanosarcina acetivorans*
28	54	gi|501693887	54	47991	0.09	Tungsten–containing formylmethanofuran dehydrogenase 2 subunit B	*Methanosphaerula palustris*
29	60	gi|499343098	52	52199	0.09	Acetyl–CoA synthase subunit beta	*Methanosarcina mazei*
30	32	gi|503663860	98	50278	0.09	Monomethylamine methyltransferase mtmB	*Methanosalsum zhilinae*
31	27	gi|501687782	105	50388	0.09	Coenzyme F420 hydrogenase subunit alpha	*Methanosphaerula palustris*
32	73	gi|499624733	48	52718	0.08	Acetyl–CoA synthase subunit beta	*Methanosarcina barkeri*
33	64	gi|503101970	50	57919	0.08	Aldehyde dehydrogenase	*Vulcanisaeta distributa*
34	97	gi|490137390	44	63808	0.07	Protein fwdA	*Methanofollis liminatans*
35	78	gi|499343351	47	59470	0.07	4Fe–4S ferredoxin	*Methanosarcina mazei*
36	59	gi|499768063	52	75918	0.06	Formate dehydrogenase subunit alpha	*Methanospirillum hungatei*
37	92	gi|500169041	44	72578	0.06	Disulfide reductase	*Methanoculleus marisnigri*

(*continued on next page*)

Table 3 (*continued*)

Cellular metabolism (27)							
44	35	gi|499329625	84	7984	1.77	Deoxyribonuclease	*Methanosarcina acetivorans*
45	22	gi|17380265	118	22967	0.73	Proteasome subunit alpha	*Methanosarcina thermophila*
46	30	gi|499342822	102	15502	0.71	Pyridoxamine 5–phosphate oxidase	*Methanosarcina mazei*
47	21	gi|6093782	120	27138	0.6	Proteasome subunit alpha	*Methanosarcina thermophila*
48	26	gi|499330993	107	9538	0.54	Effector protein	*Methanosarcina acetivorans*
49	29	gi|499345514	103	13262	0.37	4–carboxymuconolactone decarboxylase	*Methanosarcina mazei*
50	14	gi|499333059	155	14789	0.33	Cupin	*Methanosarcina acetivorans*
51	51	gi|504371269	58	16609	0.29	YbaK/prolyl–tRNA synthetase associated domain–containing protein	*Fervidicoccus fontis*
52	48	gi|499345198	61	17072	0.28	Peptidylprolyl isomerase	*Methanosarcina mazei*
53	61	gi|499168240	51	17740	0.27	Aspartate carbamoyltransferase	*Aeropyrum pernix*
54	19	gi|499626578	127	37655	0.25	NADP–dependent alcohol dehydrogenase	*Methanosarcina barkeri*
55	86	gi|499490599	45	23940	0.19	50S ribosomal protein L1	*Picrophilus torridus*
56	34	gi|511307100	87	48911	0.19	Glutamate dehydrogenase GdhA	*Methanobrevibacter sp. AbM4*
57	84	gi|495251456	45	26401	0.17	Cytochrome C	*Haladaptatus paucihalophilus*
58	15	gi|1199638	149	63804	0.14	A1AO H+ ATPase, subunit A	*Methanosarcina mazei Go1*
59	53	gi|46396470	55	32230	0.14	Pyridoxal biosynthesis lyase PdxS	*Methanosarcina acetivorans C2A*
60	67	gi|494645170	49	38444	0.12	SAM–dependent methyltransferase	*Candidatus Nitrosoarchaeum limnia*
61	90	gi|499343317	45	37687	0.12	Endonuclease	*Methanosarcina mazei*
62	76	gi|503449780	47	37850	0.12	Methylthioribose–1–phosphate isomerase	*Archaeoglobus veneficus*
63	70	gi|499316941	48	39075	0.11	Endonuclease	*Pyrobaculum aerophilum*
64	95	gi|494805725	44	39785	0.11	UDP–N–acetylglucosamine 2– epimerase	*Haloferax larsenii*
65	80	gi|493478001	47	48340	0.09	IS1341–type transposase (TCE32)	*Natrinema versiforme*
66	87	gi|505135913	45	49042	0.09	Nucleotide sugar dehydrogenase	*Natronococcus occultus*
67	82	gi|505304812	46	58596	0.08	Dihydroxyacid dehydratase	*Thermoplasmatales archaeon BRNA1*
68	66	gi|505222968	49	62963	0.07	Flagella biogenesis protein FlaI	*Natronomonas moolapensis*
69	74	gi|339756697	48	62119	0.07	Arginyl–tRNA synthetase	*Candidatus Nanosalinarum*
70	52	gi|519065234	55	63573	0.07	NADH–quinone oxidoreductase subunit C	*Halarchaeum acidiphilum*
Stress response (6)							
38	71	gi|499627762	48	23931	0.19	Superoxide dismutase	*Methanosarcina barkeri*
39	46	gi|499344808	62	24611	0.19	Superoxide dismutase	*Methanosarcina mazei*
40	94	gi|494102747	44	16407	0.29	Universal stress protein	*Methanotorris formicicus*
41	18	gi|499333494	131	16370	0.29	Universal stress protein	*Methanosarcina acetivorans*
42	31	gi|499332655	101	10442	2.27	Thioredoxin	*Methanosarcina acetivorans*
43	50	gi|499625148	58	17859	0.26	Heat–shock protein	*Methanosarcina barkeri*
Tranportor and membrane proteins (5)							
71	4	gi|499333555	419	50365	1.33	V–type ATP synthase subunit B	*Methanosarcina acetivorans*
72	6	gi|393715204	273	31948	0.94	S–Layer (Ma0829) Protein	*Methanosarcina Acetivorans*
73	36	gi|339757710	80	27518	0.17	ABC–type Fe3+–hydroxamate transport system, periplasmic	*Candidatus Nanosalina sp. J07AB43*
74	45	gi|493940982	62	39158	0.11	Basic membrane protein	*Halosimplex carlsbadense*
75	39	gi|499180850	68	43703	0.1	ABC transporter substrate–binding protein	*Archaeoglobus fulgidus*

and the inoculum, but without coal; and 3) one replicate which included the coal samples and the standard medium, but without the inoculum. For the third group, one replicate was lost during sample preparation. As shown in Table 1, a large number of proteins with confidence level of 95% ($p < 0.05$) were detected in all five samples. Using a Mascot cutoff score of 43, all observed proteins were divided into two

Table 3 (*continued*)

Function unknown (23)							
76	9	gi|499345228	198	8190	6.27	Hypothetical protein	*Methanosarcina mazei*
77	44	gi|499343248	64	10387	1.2	Hypothetical protein	*Methanosarcina mazei*
78	13	gi|499329659	160	10844	1.14	Hypothetical protein	*Methanosarcina acetivorans*
79	62	gi|490177258	51	5316	1.1	Hypothetical protein	*Methanoplanus limicola*
80	47	gi|499643138	61	15319	0.31	Hypothetical protein	*Natronomonas pharaonis*
81	96	gi|505225445	44	15467	0.31	Hypothetical protein	*Methanosarcina mazei*
82	58	gi|499466785	53	20644	0.23	Hypothetical protein	*Nanoarchaeum equitans*
83	56	gi|546143292	54	22175	0.21	Hypothetical protein	*Ferroplasma sp. Type II*
84	38	gi|499726032	69	24308	0.19	Hypothetical protein	*Methanosphaera stadtmanae*
85	83	gi|502721805	46	26613	0.17	Hypothetical protein	*Methanobrevibacter ruminantium*
86	93	gi|495717708	44	36952	0.12	Hypothetical protein	*Halorubrum californiense*
87	88	gi|501015082	40	76751	0.12	Hypothetical protein	*Methanococcus maripaludis*
88	68	gi|549635143	49	37246	0.12	Hypothetical protein	*Aeropyrum camini*
89	72	gi|500015331	48	37663	0.12	Hypothetical protein	*Methanosaeta thermophila*
90	65	gi|490728853	50	42069	0.11	GTP–binding HSR1–like protein	*Methanocaldococcus villosus*
91	85	gi|502720837	45	40956	0.11	Hypothetical protein	*Methanobrevibacter ruminantium*
92	98	gi|493181451	43	42681	0.1	Hypothetical protein	*Natrinema pellirubrum*
93	49	gi|505225764	59	55553	0.08	Hypothetical protein	*Methanosarcina mazei*
94	75	gi|495799237	47	66430	0.07	Hypothetical protein	*Halorhabdus tiamatea*
95	81	gi|499330023	46	74454	0.06	Hypothetical protein	*Methanosarcina acetivorans*
96	79	gi|499466667	47	82280	0.05	Hypothetical protein	*Nanoarchaeum equitans*
97	57	gi|499342568	54	140487	0.03	Hypothetical protein	*Methanosarcina mazei*
98	77	gi|506269996	47	206988	0.02	Hypothetical protein	*Halorhabdus utahensis*

groups: statistically valid (≥43) and statistically uncertain (<43) groups. Searching the peaklists against domain of bacteria, archaea and fungi revealed the presence of 152 ± 1.4, 96.5 ± 2.1 and 38 ± 1.4 protein families, respectively, for the first set of samples. Regarding the second set, the numbers of protein families were 240.5 ± 7.8 for bacteria, 64 ± 29.7 for archaea and 80 ± 17.0 for fungi. The third set contained 115 proteins related to bacteria, 24 proteins belonged to archaea and 19 proteins related to fungi. Comparing all three sets, there were 39 bacterial proteins, 37 from archaea and 10 fungal proteins that were present only in the first set, but not in the second and third group. Since methane was observed only in the first set of microcosms, these proteins might be strongly related to methane production. A total of 21, 7 and 2 proteins that were from bacteria, archaea and fungi, respectively, were present in all five samples. These proteins may be essential for basic cellular metabolisms.

Considering the fact that the goal of this study was to understand the pathway from coal to methane, only families of proteins in the first set of samples were analyzed. In addition, since the difference between proteins in the two replicates of set #1 was minimal, only detailed analysis of proteins in replicate 1 or sample #1 was conducted. For this sample, the total number of 151 bacterial protein families was divided into five categories (Table 2). The first group of transporter and membrane proteins included 50 protein families. Among these, one membrane protein had a high emPAI (exponentially modified protein abundance index) value of 2.37 indicating that it was abundantly produced by bacterial cells. A sum of 18 proteins was identified as ATP-binding cassette (ABC) transporters responsible for transporting sugars, amino acids and phosphonate molecules. In addition, proteins related to transporting C-4 carboxylates (malate, succinate and fumarate) and binding to starch and collagen similar compounds were also present. The fact that 33.1% of bacterial proteins were related to substrate binding and transport demonstrated that bacterial cells devoted significant amount of energy and effort trying to grab whatever was available in the microcosms.

The second category was cellular metabolism. Among 43 of this group, glutamate dehydrogenase was the most abundant with an emPAI of 1.38. This enzyme converts glutamate to oxoglutarate while releasing NH_4^+ through the reaction (Buckel, 2001). It is unknown at this stage whether the presence of five different families of this enzyme in the day-30 microcosms was related to nitrogen deficiency. The third group included 11 proteins that might be involved in substrate utilization and conversion. From the five representative enzymes: iron hydrogenase, glucokinase, rhamnulokinase, glycoside hydrolase and sulfite reductase, it could be deduced that the bacterial cells could utilize glucose, rhamnose, mixed sugars and sulfite for growth and for producing hydrogen. This group also had one family of methyl-CoM reductase (Mcr) which is the enzyme responsible for converting methyl-CoM to methane. Abundance of this protein family is fairly high with an emPAI of 0.45. Since steps and enzymes unique to the aceticlastic pathway are widely distributed in the domain of bacteria (Ferry, 2010),

Table 4
List of statistically valid proteins identical to those in fungi. Highlighted are those that are present in samples #1 and #2, but not in samples #3, #4 and #5.

Number	Family	Accession	Score	Mass (kDa)	emPAI	Protein function	Related species
Cellular metabolism (21)							
1	3	gi|115398299	82	49092	0.19	NADP–specific glutamate dehydrogenase	*Aspergillus terreus* NIH2624
2	15	gi|599100989	58	28181	0.16	N–terminal nucleophile aminohydrolase	*Punctularia strigosozonata* HHB–11173 SS5
3	16	gi|118162020	58	42274	0.11	Zinc finger transcription factor	*Cercospora nicotianae*
4	41	gi|327301591	50	39679	0.11	Translation initiation factor eIF4E3	*Trichophyton rubrum* CBS 118892
5	2	gi|242796872	83	47884	0.09	O–methyltransferase family protein	*Talaromyces stipitatus* ATCC 10500
6	7	gi|448530708	66	47925	0.09	Pmt6 protein mannosyltransferase	*Candida orthopsilosis* Co 90–125
7	22	gi|378732762	55	49128	0.09	N–acetylglucosamine–6–phosphate deacetylase	*Exophiala dermatitidis* NIH/UT8656
8	18	gi|576038998	57	54751	0.08	Putative UPF0673 membrane protein	*Chaetomium thermophilum* var. thermophilum DSM 1495
9	21	gi|528890801	56	54550	0.08	Actinin–type, actin–binding domain–containing protein	*Rozella allomycis* CSF55
10	28	gi|342321230	54	58120	0.08	Proteophosphoglycan ppg4	*Rhodotorula glutinis* ATCC 204091
11	36	gi|388851977	53	60422	0.07	Related to MSS1–mitochondrial GTPase involved in expression of COX1	*Ustilago hordei*
12	1	gi|402470658	116	87328	0.05	V–type proton ATPase catalytic subunit A	*Edhazardia aedis* USNM 41457
13	29	gi|512191814	54	89562	0.05	Vesicular–fusion protein sec18	*Ophiostoma piceae* UAMH 11346
14	30	gi|599097922	54	94013	0.05	ATP–dependent DNA helicase	*Punctularia strigosozonata* HHB–11173 SS5
15	5	gi|528891580	68	100817	0.04	ARID/BRIGHT DNA–binding domain–containing protein	*Rozella allomycis* CSF55
16	44	gi|7489931	45	115868	0.04	Major surface glycoprotein	*Pneumocystis carinii*
17	9	gi|150863758	65	142653	0.03	5–oxoprolinase	*Scheffersomyces stipitis* CBS 6054
18	38	gi|632916248	53	132725	0.03	Putative histidine kinase M232p	*Villosiclava virens*
19	20	gi|299754955	56	174814	0.02	NB–ARC	*Coprinopsis cinerea okayama* 7#130
20	32	gi|578053985	54	174310	0.02	Probable Glycogen debranching enzyme	*Zygosaccharomyces bailii* ISA1307
21	46	gi|667663171	44	229441	0.02	Pol–like protein	*Beauveria bassiana* ARSEF 2860

detection of this protein was not surprising. However, since this enzyme matched one in an uncultured bacterium (gi|167541422), no more details could be described here.

The fourth group was eight protein families that were responsive to oxidative stress. These proteins included: rubrerythrin (emPAI = 1.29) which has been proposed as a scavenger of oxygen radicals responding to oxidative stress (Lehmann et al., 1996); superoxide dismutase and peroxidase which reduces superoxide and peroxide, respectively and are described as main detoxification systems in bacteria for oxygen resistance and reduction (Zhang et al., 2006). The presence of these oxidoreductases corresponded well to transient exposure to oxygen during microcosm setup. The fifth group comprised five protein families that were related to cell movement. Flagella contribute to cell movement through chemotaxis and adhesion to host surfaces. Flagellin is the structural protein that forms the major portion of flagellar filaments (Ramos et al., 2004). The identification of five different families of flagellin proteins demonstrated that bacterial cells spent quite amount of energy in producing flagella and getting them moved to places where substrates might be available. The last group contained all of those 34 hypothetical proteins whose functions were not known at this point. These proteins might have important roles in the coal-to-methane pathway. But due to lack of studies and limited information, no specific names could be given.

Searching the generated peaklists from sample #1 against archaea domain revealed 98 protein families (Table 3). Similarly, these proteins were categorized into different groups. The first

Table 4 (*continued*)

Functions unknown (25)							
22	37	gi|320034748	53	12713	0.38	Conserved hypothetical protein	*Coccidioides posadasii* str. Silveira
23	43	gi|115401362	46	14528	0.33	Predicted protein	*Aspergillus terreus* NIH2624
24	10	gi|646310968	64	20448	0.23	Hypothetical protein PLEOSDRAFT_1100627	*Pleurotus ostreatus* PC15
25	40	gi|403416258	50	19940	0.23	Predicted protein	*Fibroporia radiculosa*
26	23	gi|475664191	54	24139	0.19	Hypothetical protein FOC4_g10015061	*Fusarium oxysporum f. sp.* cubense race 4
27	19	gi|154315511	57	25653	0.18	Predicted protein	*Botrytis cinerea* B05.10
28	12	gi|171693081	62	28886	0.16	Hypothetical protein	*Podospora anserina S mat+*
29	27	gi|552934160	54	34236	0.13	Hypothetical protein GLOINDRAFT_149240	*Rhizophagus irregularis* DAOM 181602
30	13	gi|238592521	59	42497	0.11	Hypothetical protein MPER_07428	*Moniliophthora perniciosa* FA553
31	24	gi|134080222	54	42022	0.11	Unnamed protein product	*Aspergillus niger*
32	14	gi|453085843	59	45966	0.1	Hypothetical protein SEPMUDRAFT_65229	*Sphaerulina musiva* SO2202
33	34	gi|367018766	53	45598	0.1	Hypothetical protein MYCTH_2294721	*Myceliophthora thermophila* ATCC 42464
34	6	gi|402216707	67	57230	0.08	Hypothetical protein DACRYDRAFT_112412	*Dacryopinax sp.* DJM–731 SS1
35	33	gi|599356073	54	58744	0.08	Hypothetical protein MELLADRAFT_58345	*Melampsora larici–populina* 98AG31
36	42	gi|636594601	48	54209	0.08	Hypothetical protein SETTUDRAFT_155576	*Setosphaeria turcica* Et28A
37	45	gi|525584055	45	64305	0.07	Hypothetical protein PDE_05256	*Penicillium oxalicum* 114–2
38	31	gi|170098314	54	59034	0.07	Predicted protein	*Laccaria bicolor* S238N–H82
39	11	gi|628284793	63	71212	0.06	Hypothetical protein A1O7_02672	*Cladophialophora yegresii* CBS 114405
40	35	gi|169603856	53	80074	0.05	Hypothetical protein SNOG_04936	Phaeosphaeria nodorum SN15
41	39	gi|354546729	52	83463	0.05	Hypothetical protein CPAR2_211050	Candida parapsilosis
42	26	gi|646297422	54	104454	0.04	Hypothetical protein BOTBODRAFT_171398	*Botryobasidium botryosum* FD–172 SS1
43	4	gi|646305786	71	130858	0.03	Hypothetical protein PLEOSDRAFT_1097340	*Pleurotus ostreatus* PC15
44	8	gi|528297727	65	127414	0.03	Hypothetical protein BGHDH14_bghG002813000001001	*Blumeria graminis f. sp.* hordei DH14
45	17	gi|557999973	57	167737	0.03	Hypothetical protein PSEUBRA_SCAF11g01160	*Pseudozyma brasiliensis* GHG001
46	25	gi|662500487	54	145831	0.03	Hypothetical protein M437DRAFT_88850	*Aureobasidium melanogenum* CBS 110374

group of 37 proteins or 38% of the total was tied to methane production. The most abundant protein (emPAI = 5.09) was methyl-coenzyme M reductase (Mcr) with eight different families. These proteins were close to those in *Methanosarcina mazei*, *Methanosarcina acetivorans*, *Methanoculleus marisnigri*, *Methanosaeta thermophile*, *Methanoplanus limicola* and an uncultured archaeon. This enzyme converts methyl coenzyme M and coenzyme B to methane and heterodisulfide, the last step in methane formation from CO_2, acetate and methanol (Scheller et al., 2013). The abundance of these proteins demonstrated that methane production was very active in the studied sample. Regarding the aceticlastic pathway from acetate to methane (Hanreich et al., 2012), three enzymes, acetate kinase, phosphotransacetylase and acetyl-CoA decarbonylase synthase were detected. In terms of CO_2 reduction, three out of the six enzymes dedicated for CO_2 reduction were observed. These enzymes were: formyl-methanofuran dehydrogenase (Fmd, the first enzyme in CO_2 reduction (Sakai et al., 2011)); methylene tetrahydromethanopterin (H_4MPT) dehydrogenase (Mtd; the fourth enzyme in CO_2 reduction); and methylene tetrahydromethanopterin (H_4MPT) reductase (Mer, the fifth enzyme in CO_2 reduction). In addition, four different protein families of methanol-5-hydroxybenzimidazolylcobamide methyltransferase were identified. Two proteins (gi|499331135, gi|12751300) resembled those in *M. acetivorans* and the other two (gi|499625345, gi|499627806) were similar to those in *Methanosarcina barkeri*. These proteins catalyze the transfer of a methyl group from methanol to a methanol-specific corrinoid protein (Mta) and are involved

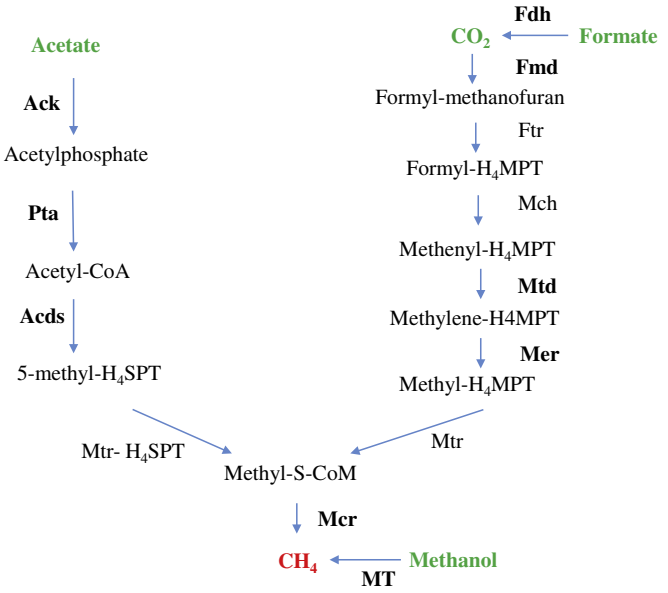

Fig. 2. A proposed pathway from acetate, CO_2, methanol and formate to methane. Proteins identified in this study were in bold. Starting compounds were in green and methane was in red. Abbreviations: Fdh, formate dehydrogenase; Fmd, formylmethanofuran dehydrogenase; Ftr, formylmethanofuran:H_4MPT formyltransferase; Mch, methenyl-H_4MPT cyclohydrolase; Mtd, F420-dependent methylene-H_4MPT dehydrogenase; Mer, methylene-H_4MPT reductase; Mtr, methyl-H_4MPT: coenzyme M methyltransferase; Mcr, methyl-coenzyme M reductase; ACK: acetate kinase; PTA: phosphotransacetylase; ACDS: acetyl CoA decarbonylase synthase; MT: methanol-5-hydroxybenzimidazolylcobamide methyltransferase; Mtr-H_4SPT: methyl-H_4SPT: coenzyme M methyltransferase. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in methanogenesis from methanol. Furthermore, two families (gi|490139462, gi|499768063) of formate dehydrogenase resembling that in *Methanofollis liminatans* and *Methanospirillum hungatei*, respectively, were detected. These proteins can oxidize formate to CO_2 (Sakai et al., 2011).

The second group was 27 protein families that were related to cellular metabolism. The most abundant one (emPAI = 1.77) was a deoxyribonuclease that degrades DNA. Other families were involved in protein synthesis and degradation, electron transport, etc. The third group included six proteins that appeared to be responding to oxidative stress. Among the six, thioredoxin had an emPAI of 2.27. This enzyme is known to reduce hydrogen peroxide and certain radicals. Similarly to bacterial strains, superoxide dismutase was also identified besides universal stress proteins. The fourth group contained five proteins that were dedicated for substrate binding and transport. These proteins included a V-type ATP synthase (emPAI = 1.33) which uses ATP hydrolysis to drive the transport of protons across a membrane; and ABC transporters with binding-substrate unknown. The fifth group was 23 hypothetical proteins whose exact functions were unclear at this stage. One protein with an emPAI of 6.27 indicated its great abundance in the microcosms. These families might represent novel proteins that were critical for the microbial community. Future research is needed to further understand functions of these unknown proteins.

A total of 46 fungal protein families were also detected in sample #1 (Table 4). These proteins were divided into two groups: cellular metabolism and hypothetical ones. The first group of 21 included proteins such as: NADP-specific glutamate dehydrogenase, transcription factors, methyltransferase, ATPase, surface glycoprotein, histidine kinase, etc. The number of proteins with unknown functions was 25. Compared to those belonging to bacteria and archaea, all of these fungal proteins were produced at low levels according to the low emPAI values.

Anaerobic fungi with large populations have been shown to colonize plant fragments in the rumen of cattle and sheep on fibrous diets (Bauchop, 1981; Orpin and Joblin, 1997) and in anaerobic digesters treating organic wastes (Schnürer and Schnürer, 2006). However, although metaproteomic studies have been conducted on plant-based feedstocks (Hanreich et al., 2013) and agricultural wastes (Heyer et al., 2013) as detailed above, no studies have reported the presence of fungal proteins. The low abundance of fungal proteins as demonstrated in this study might explain why fungal proteins have never been demonstrated before in anaerobic digesters. Adding to these low levels of proteins, the traditional way of 2-D PAGE only enables proteins with high density to be picked and identified (Abram et al., 2011; Hanreich et al., 2012, 2013). Thus, assisted by the high resolution LC/MS, this is the first study to report the presence of fungal proteins in anaerobic bioreactors converting coal to methane. This is in agreement with our microscopic observation that filamentous fungal species did exist in the microcosms. Based on DNA sequencing reported in our previous study, fungal stains came from the coal samples that we have been using and were not present in the original microbial community collected from a coal-bed methane (CBM) well (Zhang et al., 2015). If other coal samples are used in similar studies, fungal strains and related proteins may not be present. Regarding this study, although 46 proteins were related to fungi, fungal enzymes specific to coal hydrolysis, fermentation and methane production were not obviously detected in sample #1. Thus, the roles and functions of the fungal strains in the microbial community cannot be elucidated here.

3.3. Pathway from coal to methane

Over the years, several pathways based upon identified proteins have been proposed for methane formation from different substrates, for example, synthetic glucose-based wastewater (Abram et al., 2011) and beet and rye silage (Hanreich et al., 2012). With regard to coal, one pathway was proposed based on microorganisms identified through constructed clone libraries (Strąpoć et al., 2008). This pathway mainly described the coal fragmentation part and indicated that aromatic compounds, such as: polyaromatic hydrocarbons (PAHs), monoaromatic carboxylic acids and ketones were the intermediates from complex coal macromolecules. In another study, however, PAHs and ketones were not observed. Instead, single-ring aromatics, long-chain alkanes and long-chain fatty acids accumulated during the first 39 days of the 78-day study (Jones et al., 2010). This difference could be explained by examining bacterial species in the two different communities. In Strapoc's study, bacteria at the phylum level, such as Spirochaetes, Bacteroidetes, Firmicutes were identified. In Jones' study, the dominant bacteria (56%) had 99% sequence similarity to *Proteobacteria* with 43% of the clones similar to Betaproteobacteria and 13% identical to Gammaproteobacteria. Regarding bacterial populations in our tested samples, Proteobacteria, Firmicutes and Bacteroidetes were 57.2%, 33.6%, and 6.7%, respectively (Zhang et al., 2015). Thus, considering the similarities of microorganisms between reported studies and this one, it is reasonable to assume that the intermediate products from coal hydrolysis in our microcosms should be similar to those reported by the two studies. However, detailed chemical studies on identifying coal degradation products are needed to prove this assumption.

As discussed above, by taking advantage of the state-of-the-art highly sensitive LC/MS, we have identified the highest number of protein families besides novel fungal proteins. Identification of these proteins allowed us to propose a pathway from coal to methane, in particular, steps involved in methanogesis. As shown in Fig. 2, key enzymes responsible for converting acetate, CO_2, formate and methanol to methane have been identified in the studied samples. Specific to the aceticlatic pathway, one methyltransferase (gi|499329835) that was close to that in *M. acetivorans* was detected. But it is unclear whether this enzyme can catalyze the step from 5-methyl-tetrahydrosarcinapterin to methyl-CoM. Compared to other pathways proposed in the literature, which are based on either genes detected or microbes identified, the one developed from this study is the most complete. The availability

of this pathway will certainly assist future effort in optimizing methane yield from coal.

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Acknowledgments

This material is based upon work supported by the Illinois Clean Coal Institute under the Award Number of 13/4C-2 and the Department of Energy under Award Number DE-FE0024126.

References

Abram, F., Enright, A.M., O'Reilly, J., Botting, C., Collins, G., O'Flaherty, V., 2011. A metaproteomic approach gives functional insights into anaerobic digestion. J. Appl. Microbiol. 110, 1550–1560.

Ayers, W.B., 2002. Coalbed gas systems, resources, and production and a review of contrasting cases from the San Juan and powder river basins. Am. Assoc. Pet. Geol. Bull. 86, 1853–1890.

Bauchop, T., 1981. The anaerobic fungi in rumen fibre digestion. Agric. Environ. 6, 339–348.

Bonin, A.S., Boone, D.R., 2006. The order methanobacteriales. The Prokaryotes. Springer, pp. 231–243.

Buckel, W., 2001. Unusual enzymes involved in five pathways of glutamate fermentation. Appl. Microbiol. Biotechnol. 57, 263–273.

Faiz, M., Hendry, P., 2006. Significance of microbial activity in Australian coal bed methane reservoirs — a review. Bull. Can. Petrol. Geol. 54, 261–272.

Fallgren, P.H., Jin, S., Zeng, C., Ren, Z., Lu, A., Colberg, P.J.S., 2013. Comparison of coal rank for enhanced biogenic natural gas production. Int. J. Coal Geol. 115, 92–96.

Ferry, J.G., 2010. How to make a living by exhaling methane. Annu. Rev. Microbiol. 64, 453–473.

Flores, R.M., Rice, C.A., Stricker, G.D., Warden, A., Ellis, M.S., 2008. Methanogenic pathways of coal-bed gas in the Powder River Basin, United States: the geologic factor. Int. J. Coal Geol. 76, 52–75.

Green, M.S., Flanegan, K.C., Gilcrease, P.C., 2008. Characterization of a methanogenic consortium enriched from a coalbed methane well in the Powder River Basin, USA. Int. J. Coal Geol. 76, 34–45.

Hanreich, A., Heyer, R., Benndorf, D., Rapp, E., Pioch, M., Reichl, U., Klocke, M., 2012. Metaproteome analysis to determine the metabolically active part of a thermophilic microbial community producing biogas from agricultural biomass. Can. J. Microbiol. 58, 917–922.

Hanreich, A., Schimpf, U., Zakrzewski, M., Schlüter, A., Benndorf, D., Heyer, R., Rapp, E., Pühler, A., Reichl, U., Klocke, M., 2013. Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic batch fermentations indicate concerted plant carbohydrate degradation. Syst. Appl. Microbiol. 36, 330–338.

Harding, R., Czarnecki, S., Isbister, J., & Barik, S. 1993. Biogasification of low-rank coal (No. EPRI-TR-101572). Electric Power Research Inst., Palo Alto, CA (United States); ARCTECH, Inc., Chantilly, VA (United States).

Heyer, R., Kohrs, F., Benndorf, D., Rapp, E., Kausmann, R., Heiermann, M., Klocke, M., Reichl, U., 2013. Metaproteome analysis of the microbial communities in agricultural biogas plants. New Biotechnol. 30, 614–622.

Jones, E.J., Voytek, M.A., Corum, M.D., Orem, W.H., 2010. Stimulation of methane generation from nonproductive coal by addition of nutrients or a microbial consortium. Appl. Environ. Microbiol. 76, 7013–7022.

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., von Mering, C., Vorholt, J.A., 2011. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. ISME J. 6, 1378–1390.

Lehmann, Y., Meile, L., Teuber, M., 1996. Rubrerythrin from *Clostridium perfringens*: cloning of the gene, purification of the protein, and characterization of its superoxide dismutase function. J. Bacteriol. 178, 7152–7158.

Mann, K., Mann, M., 2011. In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. Proteome Sci. 9, 10.1186.

Orpin, C., Joblin, K., 1997. The rumen anaerobic fungi. The Rumen Microbial Ecosystem. Springer, pp. 140–195.

Park, C., Novak, J.T., Helm, R.F., Ahn, Y.-O., Esen, A., 2008. Evaluation of the extracellular proteins in full-scale activated sludges. Water Res. 42, 3879–3889.

Penner, T.J., Foght, J.M., Budwill, K., 2010. Microbial diversity of western Canadian subsurface coal beds and methanogenic coal enrichment cultures. Int. J. Coal Geol. 82, 81–93.

Ramos, H.C., Rumbo, M., Sirard, J.-C., 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. 12, 509–517.

Sakai, S., Takaki, Y., Shimamura, S., Sekine, M., Tajima, T., Kosugi, H., Ichikawa, N., Tasumi, E., Hiraki, A.T., Shimizu, A., 2011. Genome sequence of a mesophilic hydrogenotrophic methanogen *Methanocella paludicola*, the first cultivated representative of the order Methanocellales. PLoS ONE 6, e22898.

Scheller, S., Goenrich, M., Thauer, R.K., Jaun, B.M., 2013. Methyl-coenzyme M reductase from Methanogenic archaea: isotope effects on the formation and anaerobic oxidation of methane. J. Am. Chem. Soc. 135, 14975–14984.

Schnürer, A., Schnürer, J., 2006. Fungal survival during anaerobic digestion of organic household waste. Waste Manag. 26, 1205–1211.

Scott, A.R., Kaiser, W.R., Ayers Jr., W.B., 1994. Thermogenic and secondary biogenic gases, San Juan Basin, Colorado and New Mexico—implications for coalbed gas producibility. AAPG Bull. Am. Assoc. Pet. Geol. 78, 1186–1209.

Stokke, R., Roalkvam, I., Lanzen, A., Haflidason, H., Steen, I.H., 2012. Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. Environ. Microbiol. 14, 1333–1346.

Strąpoć, D., Picardal, F.W., Turich, C., Schaperdoth, I., Macalady, J.L., Lipp, J.S., Lin, Y.-S., Ertefai, T.F., Schubotz, F., Hinrichs, K.-U., 2008. Methane-producing microbial community in a coal bed of the Illinois Basin. Appl. Environ. Microbiol. 74, 2424–2432.

Strąpoć, D., Mastalerz, M., Dawson, K., Macalady, J., Callaghan, A.V., Wawrik, B., Turich, C., Ashby, M., 2011. Biogeochemistry of microbial coal-bed methane. Annu. Rev. Earth Planet. Sci. 39, 617–665.

Ulrich, G., Bower, S., 2008. Active methanogenesis and acetate utilization in Powder River Basin coals, United States. Int. J. Coal Geol. 76, 25–33.

Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J., Tyson, G.W., 2014. Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. Curr. Opin. Biotechnol. 27, 55–64.

Wei, M., Yu, Z., Jiang, Z., Zhang, H., 2014. Microbial diversity and biogenic methane potential of a thermogenic-gas coal mine. Int. J. Coal Geol. 134–135 (96–107).

Zhang, W., Culley, D.E., Nie, L., Brockman, F.J., 2006. DNA microarray analysis of anaerobic *Methanosarcina barkeri* reveals responses to heat shock and air exposure. J. Ind. Microbiol. Biotechnol. 33, 784–790.

Zhang, J., Liang, Y.-N., Pandey, R., Harpalani, S., 2015. Characterizing microbial communities dedicated for converting coal to methane in situ and ex situ. Int. J. Coal Geol. http://dx.doi.org/10.1016/j.coal.2015.05.001 (in press).